

```
Cloning into 'comp_4010_group'...
remote: Enumerating objects: 60, done.
remote: Counting objects: 100% (60/60), done.
remote: Compressing objects: 100% (38/38), done.
remote: Total 60 (delta 18), reused 56 (delta 14), pack-reused 0 (from 0)
Receiving objects: 100% (60/60), 282.23 KiB | 14.11 MiB/s, done.
Resolving deltas: 100% (18/18), done.

/content/comp_4010_group

Obtaining file:///content/comp_4010_group

Installing build dependencies ... done
Checking if build backend supports build_editable ... done
Getting requirements to build editable ... done
Preparing editable metadata (pyproject.toml) ... done

Requirement already satisfied: numpy in /usr/local/lib/python3.12/dist-packages (from rl-inventory==0.1.0) (2.0.2)

Requirement already satisfied: torch in /usr/local/lib/python3.12/dist-packages (from rl-inventory==0.1.0) (2.9.0+cu126)

Requirement already satisfied: gymnasium in /usr/local/lib/python3.12/dist-packages (from rl-inventory==0.1.0) (1.2.2)

Requirement already satisfied: matplotlib in /usr/local/lib/python3.12/dist-packages (from rl-inventory==0.1.0) (3.10.0)

Requirement already satisfied: pandas in /usr/local/lib/python3.12/dist-packages (from rl-inventory==0.1.0) (2.2.2)

Collecting stable-baselines3[extra] (from rl-inventory==0.1.0)
  Downloading stable_baselines3-2.7.0-py3-none-any.whl.metadata (4.8 kB)

Requirement already satisfied:云pickle>=1.2.0 in /usr/local/lib/python3.12/dist-packages (from gymnasium->rl-inventory==0.1.0) (3.1.2)
```

Requirement already satisfied: typing-extensions>=4.3.0 in /usr/local/lib/python3.12/dist-packages (from gymnasium->rl-inventory==0.1.0) (4.15.0)

Requirement already satisfied: farama-notifications>=0.0.1 in /usr/local/lib/python3.12/dist-packages (from gymnasium->rl-inventory==0.1.0) (0.0.4)

Requirement already satisfied: contourpy>=1.0.1 in /usr/local/lib/python3.12/dist-packages (from matplotlib->rl-inventory==0.1.0) (1.3.3)

Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.12/dist-packages (from matplotlib->rl-inventory==0.1.0) (0.12.1)

Requirement already satisfied: fonttools>=4.22.0 in /usr/local/lib/python3.12/dist-packages (from matplotlib->rl-inventory==0.1.0) (4.60.1)

Requirement already satisfied: kiwisolver>=1.3.1 in /usr/local/lib/python3.12/dist-packages (from matplotlib->rl-inventory==0.1.0) (1.4.9)

Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.12/dist-packages (from matplotlib->rl-inventory==0.1.0) (25.0)

Requirement already satisfied: pillow>=8 in /usr/local/lib/python3.12/dist-packages (from matplotlib->rl-inventory==0.1.0) (11.3.0)

Requirement already satisfied: pyparsing>=2.3.1 in /usr/local/lib/python3.12/dist-packages (from matplotlib->rl-inventory==0.1.0) (3.2.5)

Requirement already satisfied: python-dateutil>=2.7 in /usr/local/lib/python3.12/dist-packages (from matplotlib->rl-inventory==0.1.0) (2.9.0.post0)

Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.12/dist-packages (from pandas->rl-inventory==0.1.0) (2025.2)

Requirement already satisfied: tzdata>=2022.7 in /usr/local/lib/python3.12/dist-packages (from pandas->rl-inventory==0.1.0) (2025.2)

Requirement already satisfied: opencv-python in /usr/local/lib/python3.12/dist-packages (from stable-baselines3[extra]->rl-inventory==0.1.0) (4.12.0.88)

Requirement already satisfied: pygame in /usr/local/lib/python3.12/dist-packages (from stable-baselines3[extra]->rl-inventory==0.1.0) (2.6.1)

Requirement already satisfied: tensorboard>=2.9.1 in /usr/local/lib/python3.12/dist-packages (from stable-baselines3[extra]->rl-inventory==0.1.0) (2.19.0)

Requirement already satisfied: psutil in /usr/local/lib/python3.12/dist-packages (from stable-baselines3[extra]->rl-inventory==0.1.0) (5.9.5)

Requirement already satisfied: tqdm in /usr/local/lib/python3.12/dist-packages (from stable-baselines3[extra]->rl-inventory==0.1.0) (4.67.1)

Requirement already satisfied: rich in /usr/local/lib/python3.12/dist-packages (from stable-baselines3[extra]->rl-inventory==0.1.0) (13.9.4)

Requirement already satisfied: ale-py>=0.9.0 in /usr/local/lib/python3.12/dist-packages (from stable-baselines3[extra]->rl-inventory==0.1.0) (0.11.2)

Requirement already satisfied: filelock in /usr/local/lib/python3.12/dist-packages (from torch->rl-inventory==0.1.0) (3.20.0)

Requirement already satisfied: setuptools in /usr/local/lib/python3.12/dist-packages (from torch->rl-inventory==0.1.0) (75.2.0)

Requirement already satisfied: sympy>=1.13.3 in /usr/local/lib/python3.12/dist-packages (from torch->rl-inventory==0.1.0) (1.14.0)

Requirement already satisfied: networkx>=2.5.1 in /usr/local/lib/python3.12/dist-packages (from torch->rl-inventory==0.1.0) (3.5)

Requirement already satisfied: jinja2 in /usr/local/lib/python3.12/dist-packages (from torch->rl-inventory==0.1.0) (3.1.6)

Requirement already satisfied: fsspec>=0.8.5 in /usr/local/lib/python3.12/dist-packages (from torch->rl-inventory==0.1.0) (2025.3.0)

Requirement already satisfied: nvidia-cuda-nvrtc-cu12==12.6.77 in
/usr/local/lib/python3.12/dist-packages (from torch->rl-inventory==0.1.0) (12.6.77)

Requirement already satisfied: nvidia-cuda-runtime-cu12==12.6.77 in
/usr/local/lib/python3.12/dist-packages (from torch->rl-inventory==0.1.0) (12.6.77)

Requirement already satisfied: nvidia-cuda-cupti-cu12==12.6.80 in
/usr/local/lib/python3.12/dist-packages (from torch->rl-inventory==0.1.0) (12.6.80)

Requirement already satisfied: nvidia-cudnn-cu12==9.10.2.21 in /usr/local/lib/python3.12/dist-packages (from torch->rl-inventory==0.1.0) (9.10.2.21)

Requirement already satisfied: nvidia-cublas-cu12==12.6.4.1 in /usr/local/lib/python3.12/dist-packages (from torch->rl-inventory==0.1.0) (12.6.4.1)

Requirement already satisfied: nvidia-cufft-cu12==11.3.0.4 in /usr/local/lib/python3.12/dist-packages (from torch->rl-inventory==0.1.0) (11.3.0.4)

Requirement already satisfied: nvidia-curand-cu12==10.3.7.77 in /usr/local/lib/python3.12/dist-packages (from torch->rl-inventory==0.1.0) (10.3.7.77)

Requirement already satisfied: nvidia-cusolver-cu12==11.7.1.2 in /usr/local/lib/python3.12/dist-packages (from torch->rl-inventory==0.1.0) (11.7.1.2)

Requirement already satisfied: nvidia-cusparse-cu12==12.5.4.2 in /usr/local/lib/python3.12/dist-packages (from torch->rl-inventory==0.1.0) (12.5.4.2)

Requirement already satisfied: nvidia-cusparselt-cu12==0.7.1 in /usr/local/lib/python3.12/dist-packages (from torch->rl-inventory==0.1.0) (0.7.1)

Requirement already satisfied: nvidia-nccl-cu12==2.27.5 in /usr/local/lib/python3.12/dist-packages (from torch->rl-inventory==0.1.0) (2.27.5)

Requirement already satisfied: nvidia-nvshmem-cu12==3.3.20 in /usr/local/lib/python3.12/dist-packages (from torch->rl-inventory==0.1.0) (3.3.20)

Requirement already satisfied: nvidia-nvtx-cu12==12.6.77 in /usr/local/lib/python3.12/dist-packages (from torch->rl-inventory==0.1.0) (12.6.77)

Requirement already satisfied: nvidia-nvjitlink-cu12==12.6.85 in /usr/local/lib/python3.12/dist-packages (from torch->rl-inventory==0.1.0) (12.6.85)

Requirement already satisfied: nvidia-cufile-cu12==1.11.1.6 in /usr/local/lib/python3.12/dist-packages (from torch->rl-inventory==0.1.0) (1.11.1.6)

Requirement already satisfied: triton==3.5.0 in /usr/local/lib/python3.12/dist-packages (from torch->rl-inventory==0.1.0) (3.5.0)

Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.12/dist-packages (from python-dateutil>=2.7->matplotlib->rl-inventory==0.1.0) (1.17.0)

Requirement already satisfied: mpmath<1.4,>=1.1.0 in /usr/local/lib/python3.12/dist-packages (from sympy>=1.13.3->torch->rl-inventory==0.1.0) (1.3.0)

Requirement already satisfied: absl-py>=0.4 in /usr/local/lib/python3.12/dist-packages (from tensorboard>=2.9.1->stable-baselines3[extra]->rl-inventory==0.1.0) (1.4.0)

Requirement already satisfied: grpcio>=1.48.2 in /usr/local/lib/python3.12/dist-packages (from tensorboard>=2.9.1->stable-baselines3[extra]->rl-inventory==0.1.0) (1.76.0)

Requirement already satisfied: markdown>=2.6.8 in /usr/local/lib/python3.12/dist-packages
(from tensorflow>=2.9.1->stable-baselines3[extra]->rl-inventory==0.1.0) (3.10)

Requirement already satisfied: protobuf!=4.24.0,>=3.19.6 in /usr/local/lib/python3.12/dist-packages
(from tensorflow>=2.9.1->stable-baselines3[extra]->rl-inventory==0.1.0) (5.29.5)

Requirement already satisfied: tensorflow-data-server<0.8.0,>=0.7.0 in
/usr/local/lib/python3.12/dist-packages (from tensorflow>=2.9.1->stable-baselines3[extra]->rl-inventory==0.1.0) (0.7.2)

Requirement already satisfied: werkzeug>=1.0.1 in /usr/local/lib/python3.12/dist-packages
(from tensorflow>=2.9.1->stable-baselines3[extra]->rl-inventory==0.1.0) (3.1.3)

Requirement already satisfied: MarkupSafe>=2.0 in /usr/local/lib/python3.12/dist-packages
(from jinja2->torch->rl-inventory==0.1.0) (3.0.3)

Requirement already satisfied: markdown-it-py>=2.2.0 in /usr/local/lib/python3.12/dist-packages
(from rich->stable-baselines3[extra]->rl-inventory==0.1.0) (4.0.0)

Requirement already satisfied: pygments<3.0.0,>=2.13.0 in /usr/local/lib/python3.12/dist-packages
(from rich->stable-baselines3[extra]->rl-inventory==0.1.0) (2.19.2)

Requirement already satisfied: mdurl~=0.1 in /usr/local/lib/python3.12/dist-packages (from
markdown-it-py>=2.2.0->rich->stable-baselines3[extra]->rl-inventory==0.1.0) (0.1.2)

Downloading stable_baselines3-2.7.0-py3-none-any.whl (187 kB)

— 187.2/187.2 kB 6.9 MB/s eta 0:00:00

Building wheels for collected packages: rl-inventory

Building editable for rl-inventory (pyproject.toml) ... done

Created wheel for rl-inventory: filename=rl_inventory-0.1.0-0.editable-py3-none-any.whl
size=3153

sha256=91da148ee7bb1dd4386b11b14738d0d81e818ae6f1ba895676cbd63265690c9a

Stored in directory: /tmp/pip-ephem-wheel-cache-
zx64n2kt/wheels/35/df/06/7b2bc59eb96a3f94705fc3148b4ed8cd39f612faac5b06b58

Successfully built rl-inventory

Installing collected packages: stable-baselines3, rl-inventory

Successfully installed rl-inventory-0.1.0 stable-baselines3-2.7.0

2025-11-24 23:48:13.734306: E external/local_xla/xla/stream_executor/cuda/cuda_fft.cc:467] Unable to register cuFFT factory: Attempting to register factory for plugin cuFFT when one has already been registered

WARNING: All log messages before absl::InitializeLog() is called are written to STDERR

E0000 00:00:1764028093.753956 975 cuda_dnn.cc:8579] Unable to register cuDNN factory: Attempting to register factory for plugin cuDNN when one has already been registered

E0000 00:00:1764028093.759887 975 cuda_blas.cc:1407] Unable to register cuBLAS factory: Attempting to register factory for plugin cuBLAS when one has already been registered

W0000 00:00:1764028093.774966 975 computation_placer.cc:177] computation placer already registered. Please check linkage and avoid linking the same target more than once.

W0000 00:00:1764028093.774991 975 computation_placer.cc:177] computation placer already registered. Please check linkage and avoid linking the same target more than once.

W0000 00:00:1764028093.774996 975 computation_placer.cc:177] computation placer already registered. Please check linkage and avoid linking the same target more than once.

W0000 00:00:1764028093.775000 975 computation_placer.cc:177] computation placer already registered. Please check linkage and avoid linking the same target more than once.

2025-11-24 23:48:13.779553: I tensorflow/core/platform/cpu_feature_guard.cc:210] This TensorFlow binary is optimized to use available CPU instructions in performance-critical operations.

To enable the following instructions: AVX2 FMA, in other operations, rebuild TensorFlow with the appropriate compiler flags.

Gym has been unmaintained since 2022 and does not support NumPy 2.0 amongst other critical functionality.

Please upgrade to Gymnasium, the maintained drop-in replacement of Gym, or contact the authors of your software and request that they upgrade.

See the migration guide at https://gymnasium.farama.org/introduction/migration_guide/ for additional information.

MODELS_DIR being used:

/content/comp_4010_group/rl_inventory/agents/ppo/trained_models

ZIPs found there: []

=====

QLEARNING

=====

Force train mode: Training new model...

Auto-generated name: 20251124_234818

Training Q-Learning agent...

Episode 20, total reward = -140464.00

Episode 40, total reward = -126387.00

Episode 60, total reward = -115685.00

Episode 80, total reward = -127784.00

Episode 100, total reward = -124428.00

Episode 120, total reward = -137069.00

Episode 140, total reward = -113481.00

Episode 160, total reward = -116977.00

Episode 180, total reward = -111636.00

Episode 200, total reward = -118464.00

Episode 220, total reward = -114831.00

Episode 240, total reward = -121543.00

Episode 260, total reward = -113445.00

Episode 280, total reward = -113843.00

Episode 300, total reward = -112399.00

Episode 320, total reward = -108337.00

Episode 340, total reward = -112171.00

Episode 360, total reward = -112199.00

Episode 380, total reward = -113761.00

Episode 400, total reward = -114086.00

Episode 420, total reward = -106538.00

Episode 440, total reward = -114384.00

Episode 460, total reward = -112705.00

Episode 480, total reward = -122631.00

Episode 500, total reward = -119972.00

Episode 520, total reward = -120711.00

Episode 540, total reward = -124362.00

Saved model to:

/content/comp_4010_group/rl_inventory/agents/qlearning/trained_models/20251124_234818.pkl

Q-Learning Evaluation Report

COSTS

Avg Daily Cost: \$509.69 (+/-9.87)

Holding Cost: \$504.20

Stockout Cost: \$161715.00

Ordering Cost: \$23816.00

SERVICE LEVEL

Fill Rate: 22.2%

Stockout Rate: 94.1%

INVENTORY

Avg Inventory: 1.4 units

PERFORMANCE

Total Reward: -186035.20

=====

PPO

=====

Force train mode: Training new model...

Auto-generated name: 20251124_234845

Training PPO agent...

TRAINING PPO AGENT

Total timesteps: 365,000

Rollout length: 512

clip range: 0.2

Save location:

/content/comp_4010_group/rl_inventory/agents/ppo/trained_models/ppo_model_20251124_234845.zip

Creating PPO model...

Total timesteps: 365000

Device: cuda

/usr/local/lib/python3.12/dist-packages/stable_baselines3/common/on_policy_algorithm.py:150: UserWarning: You are trying to run PPO on the GPU, but it is primarily intended to run on the CPU when not using a CNN policy (you are using ActorCriticPolicy which should be a MlpPolicy). See <https://github.com/DLR-RM/stable-baselines3/issues/1245> for more info. You can pass

`device='cpu'` or `export CUDA_VISIBLE_DEVICES=` to force using the CPU. Note: The model will train, but the GPU utilization will be poor and the training might take longer than on CPU.

```
warnings.warn(
```

Training PPO...

Episode 1000 | Avg Reward (last 10): -1111.38

```
100% ━━━━━━━━━━━━━━━━ 365,056/365,000 [ 0:14:43 <
0:00:00 , 414 it/s ]
```

Model saved to

/content/comp_4010_group/rl_inventory/agents/ppo/trained_models/ppo_model_20251124_234845.zip

Model saved to:

/content/comp_4010_group/rl_inventory/agents/ppo/trained_models/ppo_model_20251124_234845.zip

Saved model to:

/content/comp_4010_group/rl_inventory/agents/ppo/trained_models/20251124_234845.zip

PPO Evaluation Report

COSTS

Avg Daily Cost: \$308.62 (+/-7.39)

Holding Cost: \$15285.56

Stockout Cost: \$56002.48

Ordering Cost: \$41357.62

SERVICE LEVEL

Fill Rate: 73.3%

Stockout Rate: 39.3%

INVENTORY

Avg Inventory: 41.9 units

PERFORMANCE

Total Reward: -1126.46

=====

SAC

=====

Force train mode: Training new model...

Auto-generated name: 20251125_000337

Training SAC agent...

Training SAC for 365,000

Creating SAC model...

Total timesteps: 365000

Device: cuda

Training SAC...

Episode 1000 | Avg Reward (last 10): -1080.43

100% ————— 365,000/365,000 [1:14:15
< 0:00:00 , 83 it/s]

Model saved to

/content/comp_4010_group/rl_inventory/agents/sac/trained_models/sac_inventory.zip

Saved model to:

/content/comp_4010_group/rl_inventory/agents/sac/trained_models/20251125_000337.zip

SAC Evaluation Report

COSTS

Avg Daily Cost: \$286.87 (+/-6.95)

Holding Cost: \$23851.66

Stockout Cost: \$24541.00

Ordering Cost: \$56315.89

SERVICE LEVEL

Fill Rate: 88.2%

Stockout Rate: 22.6%

INVENTORY

Avg Inventory: 65.3 units

PERFORMANCE

Total Reward: -1047.09

=====

DYNA_Q

=====

Force train mode: Training new model...

Auto-generated name: 20251125_011757

Training Dyna-Q agent...

Saved model to:

/content/comp_4010_group/rl_inventory/agents/dyna_q/trained_models/20251125_011757.pkl

Dyna-Q Evaluation Report

COSTS

Avg Daily Cost: \$331.36 (+/-14.21)

Holding Cost: \$40597.10

Stockout Cost: \$33717.00

Ordering Cost: \$46632.00

SERVICE LEVEL

Fill Rate: 83.8%

Stockout Rate: 22.2%

INVENTORY

Avg Inventory: 111.2 units

PERFORMANCE

Total Reward: -120946.10

=====

DOUBLEDQN

=====

Force train mode: Training new model...

Auto-generated name: 20251125_012040

Training Double DQN agent...

Training Double DQN agent...

State dimension: 12

Action dimension: 9

Episode 100/1000 | Avg Reward: -1546.33 | Avg Cost: 154632.53 | Avg Loss: 26.0447 | Epsilon: 0.606

Episode 200/1000 | Avg Reward: -1225.32 | Avg Cost: 122531.83 | Avg Loss: 33.6853 | Epsilon: 0.367

Episode 300/1000 | Avg Reward: -1147.94 | Avg Cost: 114794.16 | Avg Loss: 44.2352 | Epsilon: 0.222

Episode 400/1000 | Avg Reward: -1140.64 | Avg Cost: 114063.79 | Avg Loss: 47.2216 | Epsilon: 0.135

Episode 500/1000 | Avg Reward: -1125.78 | Avg Cost: 112578.14 | Avg Loss: 45.1190 | Epsilon: 0.082

Episode 600/1000 | Avg Reward: -1090.51 | Avg Cost: 109050.87 | Avg Loss: 44.8444 | Epsilon: 0.049

Episode 700/1000 | Avg Reward: -1091.79 | Avg Cost: 109178.52 | Avg Loss: 45.8885 | Epsilon: 0.030

Episode 800/1000 | Avg Reward: -1052.03 | Avg Cost: 105203.05 | Avg Loss: 42.5259 | Epsilon: 0.018

Episode 900/1000 | Avg Reward: -1077.87 | Avg Cost: 107786.53 | Avg Loss: 44.8441 | Epsilon: 0.011

Episode 1000/1000 | Avg Reward: -1096.77 | Avg Cost: 109677.38 | Avg Loss: 52.0344 | Epsilon: 0.010

Training completed!

Saved model to:
`/content/comp_4010_group/rl_inventory/agents/DoubleDQN/trained_models/20251125_012040.pt`

Double DQN Evaluation Report

COSTS

Avg Daily Cost: \$306.54 (+/-5.17)

Holding Cost: \$35812.30

Stockout Cost: \$20981.00

Ordering Cost: \$55094.00

SERVICE LEVEL

Fill Rate: 90.0%

Stockout Rate: 15.7%

INVENTORY

Avg Inventory: 98.1 units

PERFORMANCE

Total Reward: -1118.87

=====

SUMMARY COMPARISON

=====

Algorithm	Avg Cost	Fill Rate	Stockout Rate	Avg Inventory	Total Reward
SAC	286.872759	0.882185	0.225753	65.347021	-1047.085569
Double DQN	306.540548	0.899691	0.156986	98.115890	-1118.873000
PPO	308.618269	0.732902	0.392603	41.878257	-1126.456682
Dyna-Q	331.359178	0.838438	0.221918	111.224932	-120946.100000
Q-Learning	509.685479	0.222492	0.941096	1.381370	-186035.200000

Figure(1500x500)